The type I error rate for in vivo Comet assay data when the hierarchical structure is disregarded

Hansen, Merete Kjær; Kulahci, Murat

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

The type I error rate for *in vivo* Comet assay data when the hierarchical structure is disregarded

DTU Compute Technical Report-2014-09

Merete Kjær Hansen* and Murat Kulahci

Department of Applied Mathematics and Computer Science, Technical University of Denmark, DK–2800 Kgs. Lyngby, Denmark

June 15, 2014

*Corresponding author, e-mail: mkha@dtu.dk

Abstract

The Comet assay is a sensitive technique for detection of DNA strand breaks. The experimental design of *in vivo* Comet assay studies are often hierarchically structured, which should be reflected in the statistical analysis. However, the hierarchical structure sometimes seems to be disregarded, and this imposes considerable impact on the type I error rate. This study aims to demonstrate the implications that result from disregarding the hierarchical structure. Different combinations of the factor levels as they appear in a literature study give type I error rates up to 0.51 and for all combinations the type I error rate is greater than the nominal α at 0.05. Closed-form expressions based on scaled F-distributions using the Welch-Satterthwaite approximation are provided to show how the type I error rate is affected. With this study we hope to motivate researchers to be more precise regarding the exposition of the statistical methodology and to suitably account for the hierarchical structure of Comet assay data whenever present.

Contents

1 Introduction

Damage to our DNA occurs continuously due to both endogenous (e.g. metabolic processes) and exogenous (e.g. environmental agents) factors. DNA repair mechanisms are effective and constantly active, but some damages are irreparable. Accumulation of damages to the DNA may eventually become hazardous, as it may lead to unregulated cell division and tumors may evolve (Jeggo and Löbrich, 2007). The Comet assay is a rapid and sensitive technique for measuring DNA strand breaks within mammalian cells. The name of the assay originates from the images of comet-like structures that emerge due to DNA migration during electrophoresis of treated cells (Kumaravel and Jha, 2006; Hartmann et al., 2003).

A common design of the *in vivo* Comet assay entails hierarchically structured data. However, this does not seem to be accounted for in the statistical analysis. This led us to investigate the implications in terms of the type I error rate when the hierarchical structure of the data is disregarded. The type I error rate for two different hierarchical structures were assessed and it was investigated whether the the type I error rate considerably exceeded the nominal α. Closed-form expressions are provided for one of these cases.

A literature study revealed that it was not possible to determine exactly how data were analyzed due to an inadequate description. This is unfortunate since it impedes reproducibility and blurs the interpretation of the reported results. Although some researchers may analyze data properly, we find it likely that others are inspired by the insufficient description in the papers and thereby unintentionally may fail to allow properly for the nested structure.

The aim of this study is twofold. First, we aim to shed light on the insufficient description of the statistical modeling that currently characterizes some papers describing Comet assay data. Second, the implications of disregarding the hierarchical structure of data in the statistical modeling are assessed.

All results and derivations in this report assume balanced data, i.e. that the number of observations in each subgroup are the same. This is usually endavoured in Comet assay studies and it is not uncommon for designed experiments in general.

The structure of the report is as follows. Section 2 describes a common design of Comet assay studies and the resulting inherent hierarchical nature of the collected data. Section 3 presents possible statistical models for fitting raw or summarized

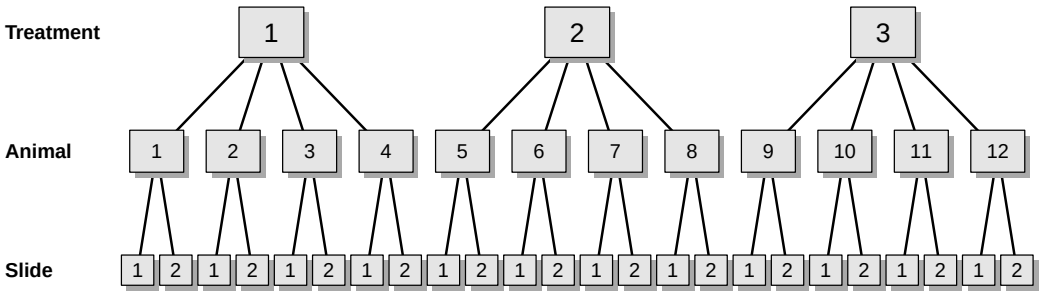

Figure 1: Outline of the design commonly used in Comet assay studies. This example shows three treatment groups, four animals per treatment and two slides per animal. For each slide a number of cells are scored, usually in the range of 50-100 cells.

Comet assay data. Section 4 describes a literature review examining the statistical analysis conducted in these published studies. Section 5 exposits the notation and relevant existing results that are used in this report. In Section 6 we look at the sampling distribution when a nested mixed-effects model is used to fit data. Section 7 provides simulated type I error rates when the hierarchical data structure is ignored in case of two different hierarchical data structures. Furthermore, closed-form expressions for the type I error rate for one of these cases is derived. Section 8 contains a discussion of the results. Some intermediate derivations are given in the Appendix.

2 Comet assay data

A common design of *in vivo* Comet assay studies is illustrated in Figure 1. Animals are randomly assigned to one of a number of different treatment groups. These treatment groups often include one negative control group, one positive control group and dose groups where increasing doses of the compound of interest are administered to the animals. For each animal there are a number of slides, in practice usually one to three slides, and from each slide a number of cells are scored.

This setup imposes a hierarchical structure of data, that is, the cells are nested within slide, that in turn is nested within animal, which again is nested within treatment. Often the interest lies in the assessment of the genotoxic effect potentially induced by the specific doses of the specific compound tested. The specific

animals used in the study is not of particular interest but merely act as representatives of the general population of that species.

50-100 cells are usually scored for each slide and the shape of the individual electrophoresed cells are very distinct. Cells can be scored both manually and automatically. One example of manual scoring is to categorise each cell in one of five categories ranging from 0 to 4 according to the shape of the cell, and a total sum is calculated for each slide or animal (Zan et al., 2013; Pesarini et al., 2013; Malta et al., 2012). Automated scoring is performed by imaging software. Popular end points are % tail DNA (percent DNA located in the Comet "tail") and the Olive tail moment, which is the product of the tail length and % tail DNA (Olive et al., 1990; Lovell and Omori, 2008). Most of the findings in the current report will be equally relevant for all types of end points assuming that they are normally distributed, possibly by transformation.

Sometimes, a summary statistic is calculated and used as response in the statistical modeling. A natural question that arises is which summary statistic to employ. Different summary statistics have been proposed, including the mean (Bright et al., 2011; Lovell et al., 1999; Wiklund and Agurell, 2003), the median (Bright et al., 2011; Lovell et al., 1999; Wiklund and Agurell, 2003; Duez et al., 2003), the 75th percentile (Lovell et al., 1999; Duez et al., 2003) and the 90th percentile (Lovell et al., 1999; Wiklund and Agurell, 2003). Also, to comply with the skewness of the within-slide distributions it has been suggested to log-transform the raw data prior to the summary calculations (Lovell and Omori, 2008). Although a few studies specifically address these issues, there is currently no consensus as to which statistic most appropriately summarizes data.

3 Statistical analysis of Comet assay data

Comet assay data can be analyzed in different ways. For some end points (e.g. % tail DNA and tail moment) data are heavily skewed and it has been suggested to model the data by means of the Weibull distribution (Ejchart and Sadlej-Sosnowska, 2003; Verde et al., 2006). In practice, it seems that only statistical methods relying on the normal distribution are used and three related statistical models valid for fitting Comet assay data are presented in the following. When data are balanced and normally distributed all three methods are equivalent. However, this requires that the statistical model matches data, i.e. if a summary statistic some-

how is calculated from the raw data this should appropriately be reflected in the model. Due to the assumption of normally distributed data it may be requisite to transform data prior to the statistical modeling.

3.1 Using raw cell scores as the response

When the raw cell scores are used as the response the hierarchical structure of data and the randomly selected animals should be properly accounted for. This can be done by employing a linear mixed-effects model with treatment as a fixed effect and animal and slide as random effects. Animal is nested within treatment and slide is nested within animal:

$$Y_{ijkl} = \mu + \tau_i + \beta_{(i)j} + \gamma_{(ij)k} + \varepsilon_{(ijk)l} \tag{1}$$

where

$$i = 1, ..., a,\ j = 1, ..., b,\ k = 1, ..., c,\ l = 1, ..., n,$$
$$\beta_{(i)j} \sim N(0, \sigma^2_\beta), \quad \gamma_{(ij)k} \sim N(0, \sigma^2_\gamma), \quad \varepsilon_{(ijk)l} \sim N(0, \sigma^2).$$

Y_{ijkl} is the $ijkl$th observation (one score for each cell) and μ and τ_i are the fixed effects for the intercept and treatment, respectively. $\beta_{(i)j}$ is the random effect of the jth animal nested within the ith treatment, $\gamma_{(ij)k}$ is the random effect of the kth slide nested within the ith treatment and jth animal and $\varepsilon_{(ij)k}$ is the within-group error. The parentheses in the subscripts indicate the nesting structure with the parent level(s) given inside the parentheses. See Montgomery (2005) for a more elaborate exposition of the linear mixed-effects model with nested effects.

3.2 Summarizing the response for each slide

Another way to analyze data is to summarize the % tail DNA distribution for each slide into a single summary statistic and use this measure in the subsequent analysis. Due to the hierarchical structure of data and the randomly selected animals a suitable analysis of the summarized data is a linear mixed-effects model with treatment as a fixed effect and animal as a random effect and with animal nested within treatment:

$$Y_{ijk} = \mu + \tau_i + \beta_{(i)j} + \varepsilon_{(ij)k} \tag{2}$$

where

$$i = 1, ..., a,\ j = 1, ..., b,\ k = 1, ..., n,$$
$$\beta_{(i)j} \sim N(0, \sigma^2_\beta), \quad \varepsilon_{(ij)k} \sim N(0, \sigma^2).$$

Y_{ijk} is the summary statistic of interest calculated for each slide and μ and τ_i are the fixed effects for the intercept and treatment, respectively. $\beta_{(i)j}$ is the random effect of the jth animal nested within the ith treatment and $\varepsilon_{(ij)k}$ is the within-group error.

3.3 Summarizing the response for each animal

A third option is to calculate a summary statistic for each animal and use this as the response. A suitable model is the fixed-effects model with treatment as a fixed effect:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \tag{3}$$

where

$$i = 1, ..., a,\ j = 1, ..., n,$$
$$\varepsilon_{ij} \sim N(0, \sigma^2).$$

Y_{ij} is the summary statistic of interest calculated for each animal, μ and τ_i are fixed effects for the intercept and treatment, respectively, and ε_{ij} is the within-group error.

4 Literature study

To investigate how data are analyzed in practice a literature study was carried out. Papers were retrieved from the search engine Web of Science with *title: in vivo* and *topic: Comet assay* from January 2012 until December 2013, which resulted in 95 papers. Of these, 47 papers conducted *in vivo* Comet assay studies with an experimental setup similar to Figure 1, and these were included in the current literature study.

Throughout the papers the execution of the experiment was well-described. This apply in particular to non-statistical aspects but also information about the number of treatment groups, number of animals per group, number of slides per animal and number of cells per slide were often clearly stated.

Regarding the statistical analysis of the Comet assay data it was in general not easy to determine how it was conducted. None of the papers defined a statistical model and no test statistics, degrees of freedom or other pointers were given. Most often it was briefly stated that data were analyzed with *one-way ANOVA* (45%), *ANOVA* (21%) or *Kruskal-Wallis test* (15%). The remaining papers predominantly used Student's t-test (also in case of more than two treatments), *Mann-Whitneys U test* or *post-hoc* tests such as Dunnett's test without preceeding use of other statistical models. None of the papers mentioned mixed models, repeated measures ANOVA, random effects, nested effects or the like.

18 papers (38%) stated "Results are expressed as mean $\pm$ SD" (or mean $\pm$ SE) or something similarly phrased. However, it was not clear how it was calculated, i.e. if these measures were calculated for each slide, for each animal etc. Also, it was not clear whether the statement was related to the tables presenting data or the statistical analysis of data. In some of these cases other summary statistics were calculated prior to the statistical analysis, i.e. in at least some cases it seems only to concern the tables summarizing data.

23 papers (49%) calculated a summary measure prior to the statistical analysis. Of these, only 15 papers (65%) clearly stated how it was done, and in these cases a summary statistic most often was calculated for each animal; that amounts to 32% of all papers that were included in the literature study. In the other 8 papers (35%) it was not possible to deduce how the summary statistic was calculated, i.e. if it was calculated per slide, per animal etc.

In 24 papers (51%) it seemed as no summary statistic was calculated prior to the statistical analysis.

The imprecise description of the statistical analysis in these papers is of a concern to us for two reasons. First, indistinctness of the methodology impedes both reproducibility and a proper interpretation of the results. Second, the combination of the lack of a calculated summary statistic and the reported statistical models that are used strongly indicates that the hierarchical structure is not properly accounted for in the statistical analysis. Although some researchers may analyze data properly, we find it likely that others are inspired by the inadequate description that implicitly suggests not to account for the hierarchical structure of data.

We performed this study to accommodate these exact concerns. By bringing these issues into focus we hope to motivate researchers to elaborate the description of

the statistical methodology. Furthermore, we wish to create awareness of the implications of ignoring potential hierarchical structure of data.

5 Notation and existing results

If $V \sim c\chi^2(\nu, \lambda)$ then V is said to follow a scaled non-central χ^2-distribution with ν degrees of freedom, scaling parameter c and non-centrality parameter λ. If $c = 1$ and $\lambda = 0$ then we say that V follows a non-scaled central χ^2-distribution. If $W \sim cF(\nu_1, \nu_2, \lambda)$ then W has a scaled non-central F-distribution with ν_1 and ν_2 degrees of freedom, scaling parameter c and non-centrality parameter λ. The cumulative distribution function of W evaluated at w is denoted $G(w; \nu_1, \nu_2, \lambda)$ when W follows a non-scaled distribution or $G_s(w; \nu_1, \nu_2, \lambda)$ if the distribution is scaled. If $W \sim F(\nu_1, \nu_2)$ then W has a non-scaled central F-distribution with the critical value $F_{\alpha;\nu_1,\nu_2}$ being the $(1-\alpha)$th quantile such that $G(F_{\alpha;\nu_1,\nu_2}; \nu_1, \nu_2) = 1 - \alpha$.

Let $X_1, X_2, ..., X_n$ be independent random variables normally distributed with expected values $E(X_1), E(X_2), \ldots, E(X_n)$ and common variance $\text{Var}(X_1) = \text{Var}(X_2) = \cdots = \text{Var}(X_n) = \text{Var}(X)$. Also, let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $E(\overline{X}) = \overline{E(X)} = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$. Then

$$V = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{\text{Var}(X)} \sim \chi^2(n-1, \lambda), \tag{4}$$

where λ is the non-centrality parameter given as

$$\lambda = \frac{\sum_{i=1}^{n} \left(E(X_i) - E(\overline{X})\right)^2}{\text{Var}(X)}$$

(Johnson et al., 1995). Furthermore, if $V_1 \sim \chi^2(\nu_1, \lambda_1)$ and $V_2 \sim \chi^2(\nu_2, \lambda_2)$ are independent random variables, then according to the reproductive property of the χ^2-distribution the sum is distributed as

$$V_1 + V_2 \sim \chi^2(\nu_1 + \nu_2, \lambda_1 + \lambda_2) \tag{5}$$

(Johnson et al., 1995; Dobson, 2002). The ratio of two independent χ^2-distributed random variables, $V_1 \sim \chi^2(\nu_1, \lambda)$ and $V_2 \sim \chi^2(\nu_2)$, each divided by its degrees of freedom follows an F-distribution with ν_1 and ν_2 degrees of freedom

$$W = \frac{V_1/\nu_1}{V_2/\nu_2} \sim F(\nu_1, \nu_2, \lambda). \tag{6}$$

with the expected value

$$E(W) = \frac{\nu_2(\nu_1 + \lambda)}{\nu_1(\nu_2 - 2)} \tag{7}$$

(Johnson et al., 1995).

Now, let $y_{ij.} = \sum_{k=1}^{n} y_{ijk}$, $y_{i..} = \sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}$, $y_{...} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}$ and let $\bar{y}_{ij.} = \frac{1}{n} y_{ij.}$, $\bar{y}_{i..} = \frac{1}{bn} y_{i..}$, $\bar{y}_{...} = \frac{1}{abn} y_{...}$. The observations y_{ijk} and the group averages $\bar{y}_{ij.}$, $\bar{y}_{i..}$ and $\bar{y}_{...}$ are realizations of the random variables Y_{ijk}, $\overline{Y}_{ij.}$, $\overline{Y}_{i..}$ and $\overline{Y}_{...}$, respectively. They are distributed as

$$\begin{aligned}
Y_{ijk} &\sim N\left(\mu + \tau_i, \sigma_\beta^2 + \sigma^2\right) \\
\overline{Y}_{ij.} &\sim N\left(\mu + \tau_i, \frac{n\sigma_\beta^2 + \sigma^2}{n}\right) \\
\overline{Y}_{i..} &\sim N\left(\mu + \tau_i, \frac{n\sigma_\beta^2 + \sigma^2}{bn}\right) \\
\overline{Y}_{...} &\sim N\left(\mu, \frac{n\sigma_\beta^2 + \sigma^2}{abn}\right)
\end{aligned} \tag{8}$$

See appendix A for details. Furthermore,

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk} - \overline{Y}_{ij.})^2 \sim \sigma^2\chi^2\left(ab(n-1)\right) \tag{9}$$

$$n\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{Y}_{ij.} - \overline{Y}_{i..})^2 \sim (\sigma_\beta^2 + \sigma^2)\chi^2\left(a(b-1)\right) \tag{10}$$

$$bn\sum_{i=1}^{a}(\overline{Y}_{i..} - \overline{Y}_{...})^2 \sim (\sigma_\beta^2 + \sigma^2)\chi^2\left(a-1, \lambda\right) \tag{11}$$

where

$$\lambda = \frac{bn \sum_{i=1}^{a} \tau_i}{n\sigma_\beta^2 + \sigma^2}$$

See appendix B for details.

6 Hierarchical models for hierarchical data

In this section we will look into the behaviour of the sampling distribution when Comet assay data summarized for each slide (i.e. as described in section 3.2) are fitted a linear mixed-effects model as defined in (2).

The hypothesis of interest is concerning equality of the different dose groups

$$\begin{aligned} H_0&: \ \tau_1 = \tau_2 = \cdots = \tau_a = 0 \\ H_1&: \ \text{at least one } \tau_i \neq 0. \end{aligned}$$

We first consider the sum of squares attributable to the treatment effect. According to (11) then

$$\frac{bn \sum_{i=1}^{a} \left(\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdots}\right)^2}{n\sigma_\beta^2 + \sigma^2} \sim \chi^2(a-1, \lambda), \tag{12}$$

where

$$\lambda = \frac{bn \sum_{i=1}^{a} \tau_i^2}{n\sigma_\beta^2 + \sigma^2}.$$

Considering the sum of squares reflecting the error component then according to (10)

$$\frac{n \sum_{i=1}^{a} \sum_{j=1}^{b} \left(\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot}\right)^2}{n\sigma_\beta^2 + \sigma^2} \sim \chi^2(a(b-1)). \tag{13}$$

As stated in (6) a ratio of two independent χ^2-distributed random variables each divided by their corresponding degrees of freedom follows an F-distribution. It

can be shown with Fisher-Cochran's theorem (Rao, 1973) that (12) and (13) are independent, hence

$$
\begin{aligned}
F_{\text{mixed}} &= \frac{\left\{ bn \sum_{i=1}^{a} \left(\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot\cdot\cdot} \right)^2 / \left(n\sigma_\beta^2 + \sigma^2 \right) \right\} / (a-1)}{\left\{ n \sum_{i=1}^{a} \sum_{j=1}^{b} \left(\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} \right)^2 / \left(n\sigma_\beta^2 + \sigma^2 \right) \right\} / (a(b-1))} \\
&= \frac{bn \sum_{i=1}^{a} \left(\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot\cdot\cdot} \right)^2 / (a-1)}{n \sum_{i=1}^{a} \sum_{j=1}^{b} \left(\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} \right)^2 / (a(b-1))} \quad \sim \quad F\left(a-1, a(b-1), \lambda \right),
\end{aligned}
\tag{14}
$$

where

$$
\lambda = \frac{bn \sum_{i=1}^{a} \tau_i^2}{n\sigma_\beta^2 + \sigma^2}. \tag{15}
$$

According to (7) the expected value of (14) is

$$
E(F_{\text{mixed}}) = \frac{a(b-1)}{(a-1)(a(b-1)-2)} \left(a - 1 + \frac{bn \sum_{i=1}^{a} \tau_i^2}{n\sigma_\beta^2 + \sigma^2} \right) \tag{16}
$$

and for sufficiently large a or b then

$$
E(F_{\text{mixed}}) \approx 1 + \frac{bn \sum_{i=1}^{a} \tau_i^2}{(a-1)(n\sigma_\beta^2 + \sigma^2)} \tag{17}
$$

which under H_0 reduces to

$$
E(F_{\text{mixed}}) \approx 1 \tag{18}
$$

7 The type I error rate - Disregarding the hierarchical structure

A type I error occurs if H_0 is rejected when it indeed is true. A type II error occurs if H_0 is not rejected although it is false (i.e. H_1 is true). A type I error

is often considered the more serious of the two and is therefore controlled more strictly. The probability of making a type I error is also called the significance level and is denoted α (Johnson et al., 2010; Hogg et al., 2005).

If some of the model assumptions are violated the actual probability of making a type I error will differ from the pre-specified significance level. Therefore, we distinguish between the former, which also is called the actual α, and the latter, which is denoted the nominal α.

From our literature study it appears as data most often are analyzed with a one-way ANOVA or Kruskal-Wallis test. However, in many cases it also seems that a suiting summary measure is not used as the response. This combination violates the assumption of independence since the observations obtained from the same animal in that case will be correlated. In the following we will investigate the implications when a one-way ANOVA is used in the analysis of hierarchically structured Comet assay data, that is, when the response is the raw cell scores as described in section 3.1 or when the response is a summary measure for each slide as described in section 3.2. The type I error rate is obtained by simulation in case of raw cell scores. Closed-form expressions for the type I error rate are provided when the response is a summary measure for each slide. The type I error rates are calculated from these expressions and are validated by simulations.

7.1 Using raw cell scores as the response

Type I error rates are in the following obtained by simulating data with a structure as depicted in Figure 1. The simulated data are subsequently analyzed by means of a one-way ANOVA, i.e. data are fitted model (3).

Table 1 shows the type I error rates for different combinations of number of treatment groups, number of animals per treatment, number of slides per animals and number of cells per animal. The levels reflect the numbers that appeared in the literature study although not all exact combinations occured. The variance components used in the simulation study were $\sigma^2_\beta = 0.08$ (animal-to-animal variation), $\sigma^2_\gamma = 0.04$ (slide-to-slide variation) and $\sigma^2 = 2.92$. These variance component equals the estimates obtained by fitting model (1) to Comet assay data obtained from an earlier study (Hansen et al., 2014). The study used % tail DNA as end point and these estimates may thus not apply to data using other end points such as the Olive tail moment or tail length. Nonetheless, the results given here can be

Table 1: Type I error rate for different combinations of number of treatment groups, animals per treatment groups, slides per animal and cells per slide. The simulated type I error rate was based on 10000 simulations for each combination (each row). The variance components used in the simulations were $\sigma^2_\beta = 0.08$ (animal-to-animal variation), $\sigma^2_\gamma = 0.04$ (slide-to-slide variation) and $\sigma^2 = 2.92$

Treatment groups	Animals per treatment	Slides per animal	Cells per slide	Simulated type I error rate
2	4	2	50	0.335
2	4	2	100	0.474
2	4	3	50	0.397
2	4	3	100	0.535
2	8	2	50	0.330
2	8	2	100	0.464
2	8	3	50	0.398
2	8	3	100	0.532
6	4	2	50	0.747
6	4	2	100	0.909
6	4	3	50	0.840
6	4	3	100	0.950
6	8	2	50	0.758
6	8	2	100	0.905
6	8	3	50	0.838
6	8	3	100	0.950

used to give an impression of the implications when the hierarchical structure is disregarded.

As seen in table 1 the type I error rate is severely inflated in all cases. The lowest type I error rate for the combinations shown here occurs when we have the lowest number of observations, namely when there is two treatment groups, four animals per treatment, two slides per animal and 50 cells per slide. Increasing the number of animals per treatment group did not affect the type I error rate much. Increasing the number of treatment groups, number of slides per animal and number of cells per slide generally resulted in increasing type I error rates. The type I error rates are between 0.335 and 0.950 and all type I error rates are thus seriously inflated. In the best case a false positive is obtained more than 3 out of 10 times whereas in the most severe case a false positive occurs more than 9 out of 10 times.

7.2 Summarizing the response for each slide

We will in the following assess the type I error rate when a summary statistic is calculated for each slide and subsequently used as the response when model (3) is fitted. First, approximate closed-form expressions are derived which aid in disclosing how the different factors affect the type I error rate. Subsequently, approximate type I error rates for different combinations of the relevant factors are calculated from the closed-form expressions and shown together with simulated type I error rates.

7.2.1 Closed-form expressions for the type I error rate

Assume that a summary measure is calculated for each slide and the fixed-effects model is employed

$$Y_{ij^*} = \mu + \tau_i + \varepsilon_{ij^*} \tag{19}$$

where $i = 1, ..., a,\ \ j^* = 1, ..., bn$ and $\varepsilon_{ij^*} \sim N(0, \sigma^{*2})$. This model typically underlies what is referred to as a one-way ANOVA. The F-statistic is calculated as

$$F_{\text{anova}} = \frac{bn \sum_{i=1}^{a} \left(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}\right)^2 / (a-1)}{\sum_{i=1}^{a} \sum_{j^*=1}^{bn} \left(Y_{ij^*} - \overline{Y}_{i\cdot}\right)^2 / (a(bn-1))} \tag{20}$$

which is expressed within the framework of model (2) as

$$F_{\text{anova}} = \frac{bn \sum_{i=1}^{a} \left(\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot\cdot\cdot}\right)^2 / (a-1)}{\sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \left(Y_{ijk} - \overline{Y}_{i\cdot\cdot}\right)^2 / (a(bn-1))} \tag{21}$$

The denominator of (21) can be rewritten as

$$\left\{ n \sum_{i=1}^{a} \sum_{j=1}^{b} \left(\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot}\right)^2 + \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \left(Y_{ijk} - \overline{Y}_{ij\cdot}\right)^2 \right\} / \left(a(b-1) + ab(n-1)\right) \tag{22}$$

implying that sum of squares and the degrees of freedom in the denominator is attributable both to the animal and the error part.

A nice feature of the F_{mixed}-statistic given in (14) is that the sum of squares in the numerator and denominator both follow χ^2-distributions that are scaled by $n\sigma_\beta^2 + \sigma^2$, that is, they cancel out and the ratio follows a standard F-distribution. This is not the case for F_{anova}-statistic in (21) as the sum of squares follow χ^2-distributions that are scaled differently. The sum of squares in the numerator is distributed as

$$bn \sum_{i=1}^{a} \left(\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot\cdot\cdot}\right)^2 \sim \left(n\sigma_\beta^2 + \sigma^2\right) \chi^2 \left(a-1, \lambda\right), \tag{23}$$

where λ is given in (15). Since Y_{ijk} are not independent the denominator of (21) does not follow the usual χ^2-distribution (see Box (1954) for details). However, looking separately at the two terms in the numerator of (22) gives

$$n \sum_{i=1}^{a} \sum_{j=1}^{b} \left(\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot}\right)^2 \sim \left(n\sigma_\beta^2 + \sigma^2\right) \chi^2(a(b-1)), \tag{24}$$

and

$$\sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \left(Y_{ijk} - \overline{Y}_{ij\cdot}\right)^2 \sim \sigma^2 \chi^2(ab(n-1)), \tag{25}$$

that is, $\sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \left(Y_{ijk} - \overline{Y}_{i\cdot\cdot}\right)^2$ is a linear combination of independent χ^2-distributed random variables. An approximate distribution is obtained using the rationale of the Welch-Satterthwaite approximation (Welch, 1938; Satterthwaite, 1941; Box, 1954). The sum of squares is approximated by a scaled χ^2-distribution

$$\sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \left(Y_{ijk} - \overline{Y}_{i\cdot\cdot}\right)^2 \dot{\sim} c\chi^2(\nu) \tag{26}$$

where c and ν are found by matching the first two moments (see appendix C). Thus,

$$\begin{aligned} c &= \frac{a(b-1)(n\sigma_\beta^2 + \sigma^2)^2 + ab(n-1)(\sigma^2)^2}{a(b-1)(n\sigma_\beta^2 + \sigma^2) + ab(n-1)\sigma^2} \\ &= \frac{(b-1)(n\sigma_\beta^2 + \sigma^2)^2 + b(n-1)(\sigma^2)^2}{(b-1)(n\sigma_\beta^2 + \sigma^2) + b(n-1)\sigma^2} \end{aligned} \tag{27}$$

and

$$\nu = \frac{(a(b-1)(n\sigma_\beta^2+\sigma^2)+ab(n-1)\sigma^2)^2}{a(b-1)(n\sigma_\beta^2+\sigma^2)^2+ab(n-1)(\sigma^2)^2}$$

$$= \frac{a((b-1)(n\sigma_\beta^2+\sigma^2)+b(n-1)\sigma^2)^2}{(b-1)(n\sigma_\beta^2+\sigma^2)^2+b(n-1)(\sigma^2)^2} \tag{28}$$

where ν is known as the effective degrees of freedom (Satterthwaite, 1941). As previously mentioned a ratio of χ^2-distributed random variables each divided by its degrees of freedom are F-distributed. However, the sum of squares in the denominator of (21) is not divided by its effective degrees of freedom ν but by $a(bn-1)$, so that

$$F_{\text{anova}} = \frac{bn\sum_{i=1}^{a}\left(\overline{Y}_{i\cdot\cdot}-\overline{Y}_{\cdot\cdot\cdot}\right)^2/(a-1)}{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk}-\overline{Y}_{i\cdot\cdot}\right)^2/\nu} \quad \cdot \quad \frac{a(bn-1)}{\nu} \tag{29}$$

In addition, adjusting for the distinct scaling of the distributions of the numerator (scaled by $n\sigma_\beta^2+\sigma^2$) and denominator (scaled by c) gives an approximate distribution of F_{anova}

$$F_{\text{anova}} \dot{\sim} \frac{a(bn-1)}{\nu}\frac{n\sigma_\beta^2+\sigma^2}{c}F(a-1,\nu,\lambda), \tag{30}$$

and inserting ν and c gives

$$F_{\text{anova}} \dot{\sim} \xi\, F(a-1,\nu,\lambda), \tag{31}$$

where

$$\xi = \frac{(bn-1)(n\sigma_\beta^2+\sigma^2)}{(b-1)(n\sigma_\beta^2+\sigma^2)+b(n-1)\sigma^2}. \tag{32}$$

According to (7) then the expected value of F_{anova} becomes

$$E(F_{\text{anova}}) \approx \xi\,\frac{\nu}{(a-1)(\nu-2)}\left(a-1+\frac{bn\sum_{i=1}^{a}\tau_i^2}{n\sigma_\beta^2+\sigma^2}\right) \tag{33}$$

and for sufficiently large ν

$$E(F_{\text{anova}}) \approx \xi \left(1 + \frac{bn \sum_{i=1}^{a} \tau_i^2}{(a-1)(n\sigma_\beta^2 + \sigma^2)}\right) \tag{34}$$

which under H_0 reduces to

$$E(F_{\text{anova}}) \approx \xi. \tag{35}$$

When $\sigma_\beta^2 = 0$ then $\xi = 1$ and under H_0 then $E(F_{\text{anova}}) \approx 1$. For $\sigma_\beta^2 > 0$ then $\xi > 1$ implying that $E(F_{\text{anova}}) > 1$.

In practice, when data are analyzed with a one-way ANOVA the observed F_{anova}-statistic is (incorrectly) compared to a critical value obtained from an unscaled F-distribution, $F_{\alpha;a-1,a(bn-1)}$. The approximate type I error rate is found as

$$\text{Type I error rate} = 1 - G_s(F_{\alpha;a-1,a(bn-1)}; a-1, \nu), \tag{36}$$

where G_s refers to the scaled cumulative distribution function of F_{anova} given in (31) with $\lambda = 0$, since the type I error rate is defined under H_0.

Multiplying the scaled F-distribution by ξ^{-1} is a monotonic transformation (i.e. it preserves the order of the quantiles), hence the type I error rate can also be calculated as

$$\text{Type I error rate} \approx 1 - G\left(\xi^{-1} F_{\alpha;a-1,a(bn-1)}; a-1, \nu\right) \tag{37}$$

and the type I error rate can be found by means of a non-scaled F-distribution, which is readily available in most statistical software.

The type I error rate can also be expressed in terms of the variance components

$$\sigma^2_{\text{ratio}} = \frac{\sigma_\beta^2}{\sigma^2} \tag{38}$$

The effective degrees of freedom and the scaling factor is then found as

$$\begin{aligned} \nu &= \frac{a((b-1)(n\sigma_\beta^2 + \sigma^2) + b(n-1)\sigma^2)^2}{(b-1)(n\sigma_\beta^2 + \sigma^2)^2 + b(n-1)(\sigma^2)^2} \cdot \frac{(\sigma^{-2})^2}{(\sigma^{-2})^2} \\ &= \frac{a((b-1)(n\sigma^2_{\text{ratio}} + 1) + b(n-1))^2}{(b-1)(n\sigma^2_{\text{ratio}} + 1)^2 + b(n-1)} \end{aligned} \tag{39}$$

and

$$\begin{aligned}\xi &= \frac{(bn-1)(n\sigma_\beta^2+\sigma^2)}{(b-1)(n\sigma_\beta^2+\sigma^2)+b(n-1)\sigma^2} \quad \cdot \quad \frac{\sigma^{-2}}{\sigma^{-2}} \\ &= \frac{(bn-1)(n\sigma_{\text{ratio}}^2+1)}{(b-1)(n\sigma_{\text{ratio}}^2+1)+b(n-1)}. \end{aligned} \tag{40}$$

implying that the distribution of the F_{anova}-statistic and hence the type I error rate is influenced by the relative magnitudes of σ_β^2 and σ^2.

The type I error rate in special cases

In the special case where $n = 1$, that is, there is one slide per animal, then

$$\nu = a(b-1) \tag{41}$$

and

$$\gamma = 1 \tag{42}$$

For $\sigma_\beta^2 = 0$ then

$$\nu = a(bn-1) \tag{43}$$

and

$$\gamma = 1 \tag{44}$$

In both cases the approximate distribution in (31) becomes the ususal (appropriate) F-distribution and the type I error rate in (37) becomes α. This is what we expect since the hierarchical structure of the data in these cases will vanish so that model (19) becomes a suitable choice.

7.2.2 Results

Table 2 summarizes the type I error rate for different combinations of treatment groups, animals per treatment, slides per animal and ratios of the variance components. The levels of the first three factors, i.e. treatment groups, animals and slides were selected among actual levels identified in the literature study, although not every combination of the three factors occurred. From an earlier study (Hansen et al., 2014), where % tail DNA was used as an end point, $\hat{\sigma}_{\text{ratio}}^2 = 0.9 \approx 1$ and

Table 2: Type I error rate for different combinations of number of treatment groups (a), animals per treatment groups (b), slides per animal (n) and variance ratio $\sigma^2_{\text{ratio}} = \frac{\sigma^2_\beta}{\sigma^2}$. The effective denominator degrees of freedom ν, the scaling parameter ξ and the approximate $E(F)$ was calculated from (39), (40) and (33), respectively. The simulated type I error rate was based on 10000 simulations for each combination (each row) and the approximate type I error rate was found from (37). All approximate type I error rates were covered by the 95% confidence intervals for the simulated type I error rates except for two cases marked by asterisks. Type I error rates greater then 0.20 are marked in bold.

Treatment groups	Animals per treatment	Slides per animal	σ^2_{ratio}	Den DF $a(bn-1)$	Den DF ν	ξ	Approximate $E(F)$	Simulated type I error rate	Approximate type I error rate
2	4	2	0.5	14	12.50	1.40	1.67	0.094	0.094
2	4	2	1.0	14	10.90	1.61	1.98	0.118	0.120
2	4	2	2.0	14	9.14	1.84	2.36	0.150	0.148
2	4	3	0.5	22	17.96	1.77	2.00	0.139	0.137
2	4	3	1.0	22	14.29	2.20	2.56	0.186	0.183
2	4	3	2.0	22	10.85	2.65	3.25	**0.227**	**0.230**
2	8	2	0.5	30	26.89	1.36	1.47	0.090	0.092
2	8	2	1.0	30	23.69	1.55	1.70	0.120	0.114
2	8	2	2.0	30	20.21	1.74	1.94	0.140	0.138
2	8	3	0.5	46	37.56	1.72	1.81	0.131	0.133
2	8	3	1.0	46	30.25	2.09	2.24	0.178	0.174
2	8	3	2.0	46	23.54	2.48	2.71	**0.223**	**0.213***
6	4	2	0.5	42	37.50	1.40	1.48	0.147	0.149
6	4	2	1.0	42	32.71	1.61	1.72	**0.212**	**0.214**
6	4	2	2.0	42	27.42	1.84	1.99	**0.283**	**0.284**
6	4	3	0.5	66	53.89	1.77	1.84	**0.268**	**0.267**
6	4	3	1.0	66	42.86	2.20	2.31	**0.386**	**0.390**
6	4	3	2.0	66	32.55	2.65	2.83	**0.509**	**0.501**
6	8	2	0.5	90	80.67	1.36	1.40	0.149	0.144
6	8	2	1.0	90	71.07	1.55	1.60	0.194	**0.203***
6	8	2	2.0	90	60.62	1.74	1.80	**0.271**	**0.265**
6	8	3	0.5	138	112.69	1.72	1.75	**0.256**	**0.257**
6	8	3	1.0	138	90.75	2.09	2.14	**0.374**	**0.371**
6	8	3	2.0	138	70.61	2.48	2.55	**0.477**	**0.473**

the levels of the ratio were selected as 0.5, 1 and 2 times this approximate estimate.

The assumed denominator degrees of freedom was calculated as $a(bn-1)$ as this is used when data are fitted model (19). The effective degrees of freedom ν, the scaling factor ξ and the approximate $E(F)$ was calculated from (39), (40) and (33), respectively. The simulated type I error rate was obtained by simulating data structured as in Figure 1, and for each combination (each row in Table 2) 10000 simulations were conducted. The approximate type I error rates were calculated from (37). Throughout the nominal α was 0.05.

In all cases the assumed denominator degrees of freedom were greater than the effective denominator degrees of freedom, ν, and furthermore $\xi > 1$. This imposes additional skewness to the F-distribution implying a heavier right tail as seen in Figure 2, which illustrates the F-distributions for six treatment groups, four animals, three slides and $\sigma^2_{\text{ratio}} = 1$.

Increasing the number of treatment groups enhanced the type I error rate considerably. The same was in evidence when the number of slides per animal were increased. Interestingly, the number of animals per treatment group did not affect the type I error rate noticeably. Increasing σ^2_{ratio} (increasing σ^2_β relative to σ^2) in general increased the type I error rate. All cases resulted in a type I error rate greater than the nominal α at 0.05. Most combinations gave type I error rates greater than 0.10 and almost half resulted in type I error rates greater than 0.20.

The validity of the approximate type I error rates was assessed by making an informal comparison to the simulated type I error rates. To quantify the simulation uncertainty the standard errors were calculated as $\text{se}(\hat{p}) = \sqrt{\hat{p}(1-\hat{p})/n}$ and Wald based 95% confidence intervals (CI) were obtained (not shown). The simulated type I error rates were between 0.090 and 0.509, hence the standard errors were between 0.003 and 0.005. In all but two cases the approximate type I error rates were covered by the CI for the simulated type I error rates. This agrees with the expectation of 1 to 2 values falling outside the CI given the number of comparisons and the confidence level. The two cases not covered by the CI are marked with asterisks in Table 2. A 99% CI for the simulated type I error rates covered all the approximate type I error rates.

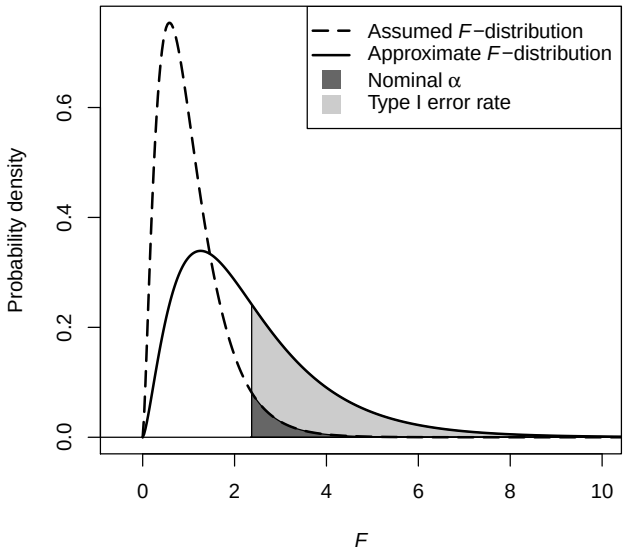

Figure 2: The F-distributions in case of six treatment groups, four animals per treatment, three slides per animal and $\sigma^2_{\text{ratio}} = 1$. The assumed F-distribution refers to the distribution from which the critical value is obtained. The approximate F-distribution is the distribution of F_{anova} as defined in (31). The approximate F-distribution has a heavier right tail implying that the type I error rate is greater than the nominal α at 0.05.

8 Discussion

This study aimed at addressing potential issues concerning the analysis of Comet assay data. First, from the literature study it was not possible to deduce exactly how data were analyzed, which impedes reproducibility and blurs the interpretation of the reported results. Even if some researchers analyze data properly, we find it likely that others (e.g. new researchers in the field) may be inspired by the insufficient description of the statistical modeling in the papers and thereby may fail to allow properly for the nested structure. Second, as we suspect that the nested structure in data is not accounted for in the statistical model we investigated the implications in terms of the type I error rate. Approximate formulas for one likely case were derived to examine in which way the type I error rate was affected.

Type I error rates for different combinations of the factors as they appeared in the literature study demonstrated that the inflation is in fact non-trivial. When the

cell scores are used as the response all type I error rates examined in the current study were severely inflated yielding type I error rates as high as 0.950. These results were seen for combination of factors as they appeared in the literatury study and we therefore consider these results likely to occur in practice. Interestingly, the variance components reflecting the animal and slide variation were relatively small compared to the residual variation, i.e. the ratios were $\frac{\sigma^2_\beta}{\sigma^2} = 0.026$ and $\frac{\sigma^2_\gamma}{\sigma^2} = 0.013$, respectively. Even so, the results show that the hypothesis test yields completely unreliable results from which erroneous inferences are made. This means that even factors that contribute with variation that seem negligible can have a huge impact on the results. One reason may be the high number of scored cells, which often in practice is 50 or 100 cells per slide.

Closed-form expressions were derived for the case where a summary statistic is calculated for each slide and they showed that the actual sampling distribution approximately follows a scaled F-distribution. Both the number of treatment groups, animals per treatment, slides per animal, the variance ratio $\sigma^2_{\text{ratio}} = \frac{\sigma^2_\beta}{\sigma^2}$ and the significance level, α, influences the shape of this distribution and hence the type I error rate. For the cases shown here the approximate type I error rates were between 0.094 and 0.501, and for all combinations they were greater than the nominal α at 0.05. Almost half of the cases resulted in type I error rates greater than 0.20. In practice, the number of animals did not seem to have a noticeable effect on the type I error rate but all other factors that appeared in the closed-form expressions affected the type I error rate appreciably.

Our objective was to illustrate the implications in a simple manner with the hope of motivating researchers within the field to reconsider the statistical modeling. As the design considered here is widespread across various scientific areas we believe that the results may be equally relevant to researchers in other fields.

Appendices

A Expectation and variance of Y_{ijk}, $\overline{Y}_{ij\cdot}$, $\overline{Y}_{i\cdot\cdot}$ and $\overline{Y}_{\cdot\cdot\cdot}$

In the following the expectation and variance of Y_{ijk}, $\overline{Y}_{ij\cdot}$, $\overline{Y}_{i\cdot\cdot}$ and $\overline{Y}_{\cdot\cdot\cdot}$ is derived from model (2). All terms in the model are assumed to be independent and the following results are used:

Let $X_1, X_2, ..., X_n$ be random variables and let $T = \sum_{i=1}^{n} a_i X_i$. Then

$$E(T) = \sum_{i=1}^{n} a_i E(X_i) \tag{45}$$

and if $X_1, X_2, ..., X_n$ are independent then

$$\mathrm{Var}(T) = \sum_{i=1}^{n} a_i^2 \mathrm{Var}(X_i) \tag{46}$$

(Hogg et al., 2005)

Expectation and variance of Y_{ijk}

Given that

$$Y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{(ij)k} \tag{47}$$

then

$$\begin{aligned} E(Y_{ijk}) &= E(\mu) + E(\tau_i) + E\left(\beta_{j(i)}\right) + E\left(\varepsilon_{(ij)k}\right) \\ &= \mu + \tau_i \end{aligned} \tag{48}$$

and

$$\begin{aligned} \mathrm{Var}(Y_{ijk}) &= \mathrm{Var}(\mu) + \mathrm{Var}(\tau_i) + \mathrm{Var}\left(\beta_{j(i)}\right) + \mathrm{Var}\left(\varepsilon_{(ij)k}\right) \\ &= \sigma_\beta^2 + \sigma^2 \end{aligned} \tag{49}$$

Expectation and variance of $\overline{Y}_{ij.}$

The group mean $\overline{Y}_{ij.}$ is obtained as

$$\begin{aligned}\overline{Y}_{ij.} &= \frac{1}{n}\sum_{k=1}^{n} Y_{ijk} \\ &= \frac{1}{n}\sum_{k=1}^{n} \left(\mu + \tau_i + \beta_{j(i)} + \varepsilon_{(ij)k}\right) \\ &= \mu + \tau_i + \beta_{j(i)} + \frac{1}{n}\sum_{k=1}^{n} \varepsilon_{(ij)k}. \end{aligned} \tag{50}$$

Then

$$\begin{aligned}E(\overline{Y}_{ij.}) &= E(\mu) + E(\tau_i) + E\left(\beta_{j(i)}\right) + E\left(\frac{1}{n}\sum_{k=1}^{n} \varepsilon_{(ij)k}\right) \\ &= \mu + \tau_i \end{aligned} \tag{51}$$

and

$$\begin{aligned}\mathsf{Var}(\overline{Y}_{ij.}) &= \mathsf{Var}(\mu) + \mathsf{Var}(\tau_i) + \mathsf{Var}\left(\beta_{j(i)}\right) + \mathsf{Var}\left(\frac{1}{n}\sum_{k=1}^{n} \varepsilon_{(ij)k}\right) \\ &= \sigma^2_\beta + \frac{\sigma^2}{n} \\ &= \frac{n\sigma^2_\beta + \sigma^2}{n} \end{aligned} \tag{52}$$

Expectation and variance of $\overline{Y}_{i\cdot\cdot}$

The group mean $\overline{Y}_{i\cdot\cdot}$ is obtained as

$$\begin{aligned}
\overline{Y}_{i\cdot\cdot} &= \frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n} Y_{ijk} \\
&= \frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n} \left(\mu + \tau_i + \beta_{j(i)} + \varepsilon_{(ij)k}\right) \\
&= \mu + \tau_i + \frac{1}{b}\sum_{j=1}^{b}\beta_{j(i)} + \frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n}\varepsilon_{(ij)k}. \qquad (53)
\end{aligned}$$

Then

$$\begin{aligned}
E(\overline{Y}_{i\cdot\cdot}) &= E(\mu) + E(\tau_i) + E\left(\frac{1}{b}\sum_{j=1}^{b}\beta_{j(i)}\right) + E\left(\frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n}\varepsilon_{(ij)k}\right) \\
&= \mu + \tau_i \qquad (54)
\end{aligned}$$

and

$$\begin{aligned}
\mathsf{Var}(\overline{Y}_{i\cdot\cdot}) &= \mathsf{Var}(\mu) + \mathsf{Var}(\tau_i) + \mathsf{Var}\left(\frac{1}{b}\sum_{j=1}^{b}\beta_{j(i)}\right) + \mathsf{Var}\left(\frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n}\varepsilon_{(ij)k}\right) \\
&= \frac{\sigma_\beta^2}{b} + \frac{\sigma^2}{bn} \\
&= \frac{n\sigma_\beta^2 + \sigma^2}{bn} \qquad (55)
\end{aligned}$$

Expectation and variance of $\overline{Y}_{...}$

The group mean $\overline{Y}_{...}$ is obtained as

$$\begin{aligned}
\overline{Y}_{...} &= \frac{1}{abn}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} Y_{ijk} \\
&= \frac{1}{abn}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} \left(\mu + \tau_i + \beta_{j(i)} + \varepsilon_{(ij)k}\right) \\
&= \mu + \frac{1}{a}\sum_{i=1}^{a}\tau_i + \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\beta_{j(i)} + \frac{1}{abn}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\varepsilon_{(ij)k}. \qquad (56)
\end{aligned}$$

Then

$$\begin{aligned}
E(\overline{Y}_{...}) &= E(\mu) + E\left(\frac{1}{a}\sum_{i=1}^{a}\tau_i\right) + E\left(\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\beta_{j(i)}\right) \\
&\quad + E\left(\frac{1}{abn}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\varepsilon_{(ij)k}\right) \\
&= \mu \qquad (57)
\end{aligned}$$

and

$$\begin{aligned}
\mathsf{Var}(\overline{Y}_{...}) &= \mathsf{Var}(\mu) + \mathsf{Var}\left(\frac{1}{a}\sum_{i=1}^{a}\tau_i\right) + \mathsf{Var}\left(\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\beta_{j(i)}\right) \\
&\quad + \mathsf{Var}\left(\frac{1}{abn}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\varepsilon_{(ij)k}\right) \\
&= \frac{\sigma_\beta^2}{ab} + \frac{\sigma^2}{abn} \\
&= = \frac{n\sigma_\beta^2 + \sigma^2}{abn} \qquad (58)
\end{aligned}$$

B Distribution of the sum of squares

In the following the distributions of the relevant sum of squares that appear in the F_{anova}-statistic presented in section 6 are derived. The results are based on the

definition of model (2) and the results obtained in appendix A.

Distribution of $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk}-\overline{Y}_{ij\cdot})^2$

According to (2) and (50) then

$$\begin{aligned} Y_{ijk}-\overline{Y}_{ij\cdot} &= \mu+\tau_i+\beta_{(i)j}+\varepsilon_{(ij)k}-\left(\mu+\tau_i+\beta_{(i)j}+\frac{1}{n}\sum_{k=1}^{n}\varepsilon_{(ij)k}\right) \\ &= \varepsilon_{(ij)k}-\frac{1}{n}\sum_{k=1}^{n}\varepsilon_{(ij)k} \end{aligned} \tag{59}$$

Since $\varepsilon_{(ij)k}\sim N(0,\sigma^2)$ and $\frac{1}{n}\sum_{k=1}^{n}\varepsilon_{(ij)k}\sim N\left(0,\frac{\sigma^2}{n}\right)$ then

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(\varepsilon_{(ij)k}-\frac{1}{n}\sum_{k=1}^{n}\varepsilon_{(ij)k}\right)^2\sim\sigma^2\chi^2\left(ab(n-1)\right) \tag{60}$$

hence

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk}-\overline{Y}_{ij\cdot}\right)^2\sim\sigma^2\chi^2\left(ab(n-1)\right) \tag{61}$$

Distribution of $n\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{Y}_{ij\cdot}-\overline{Y}_{i\cdot\cdot})^2$

According to (50) and (53) then

$$\begin{aligned} Y_{ij\cdot}-\overline{Y}_{i\cdot\cdot} &= \mu+\tau_i+\beta_{(i)j}+\frac{1}{n}\sum_{k=1}^{n}\varepsilon_{(ij)k} \\ &\quad-\left(\mu+\tau_i+\frac{1}{b}\sum_{j=1}^{b}\beta_{(i)j}+\frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n}\varepsilon_{(ij)k}\right) \\ &= \beta_{(i)j}+\frac{1}{n}\sum_{k=1}^{n}\varepsilon_{(ij)k}-\frac{1}{b}\sum_{j=1}^{b}\left(\beta_{(i)j}+\frac{1}{n}\sum_{k=1}^{n}\varepsilon_{(ij)k}\right) \end{aligned} \tag{62}$$

where the last term is seen to an average of the first two terms. Also,

$$\beta_{(i)j} + \frac{1}{n}\sum_{k=1}^{n}\varepsilon_{(ij)k} \sim N\left(0, \frac{n\sigma_\beta^2 + \sigma^2}{n}\right), \tag{63}$$

and

$$\frac{1}{b}\sum_{j=1}^{b}\left(\beta_{(i)j} + \frac{1}{n}\sum_{k=1}^{n}\varepsilon_{(ij)k}\right) \sim N\left(0, \frac{n\sigma_\beta^2 + \sigma^2}{bn}\right) \tag{64}$$

so that

$$n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot}\right)^2 \sim (n\sigma_\beta^2 + \sigma^2)\chi^2\left(a(b-1)\right) \tag{65}$$

Distribution of $bn\sum_{i=1}^{a}(\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdots})^2$

According to (53) and (56) then

$$\begin{aligned} Y_{i\cdot\cdot} - \overline{Y}_{\cdots} &= \mu + \tau_i + \frac{1}{b}\sum_{j=1}^{b}\beta_{(i)j} + \frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n}\varepsilon_{(ij)k} \\ &\quad - \left(\mu + \frac{1}{a}\sum_{i=1}^{a}\tau_i + \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\beta_{(i)j} + \frac{1}{abn}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\varepsilon_{(ij)k}\right) \\ &= \tau_i + \frac{1}{b}\sum_{j=1}^{b}\beta_{(i)j} + \frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n}\varepsilon_{(ij)k} \\ &\quad - \frac{1}{a}\sum_{i=1}^{a}\left(\tau_i + \frac{1}{b}\sum_{j=1}^{b}\beta_{(i)j} + \frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n}\varepsilon_{(ij)k}\right) \end{aligned} \tag{66}$$

where the last term is seen to be an average of the first three terms. Also,

$$\tau_i + \frac{1}{b}\sum_{j=1}^{b}\beta_{(i)j} + \frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n}\varepsilon_{(ij)k} \sim N\left(\tau_i, \frac{n\sigma_\beta^2 + \sigma^2}{bn}\right), \tag{67}$$

and

$$\frac{1}{a}\sum_{i=1}^{a}\left(\tau_i + \frac{1}{b}\sum_{j=1}^{b}\beta_{(i)j} + \frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n}\varepsilon_{(ij)k}\right) \sim N\left(0, \frac{n\sigma_\beta^2 + \sigma^2}{abn}\right), \tag{68}$$

so that

$$bn\sum_{i=1}^{a}\left(\overline{Y}_{i\cdot\cdot}-\overline{Y}_{\cdot\cdot\cdot}\right)^2\sim(n\sigma_\beta^2+\sigma^2)\chi^2\left(a-1,\lambda\right), \tag{69}$$

where

$$\lambda=\frac{bn\sum_{i=1}^{n}\tau_i^2}{n\sigma_\beta^2+\sigma^2} \tag{70}$$

C Approximate distribution of a linear combination of χ^2 variates

The sum of squares in the denominator of (21) can be partitioned as

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk}-\overline{Y}_{i\cdot\cdot}\right)^2=n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{Y}_{ij\cdot}-\overline{Y}_{i\cdot\cdot}\right)^2+\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk}-\overline{Y}_{ij\cdot}\right)^2 \tag{71}$$

which can be expressed as

$$V=V_1+V_2, \tag{72}$$

where $V_i\sim c_i\chi^2(\nu_i), i=1,...,2$. An exact distribution of V is given in Box (1954) and Satterthwaite (1941). However, a more accessible representation can be accomplished by means of the Welch-Satterthwaite approach, where the distribution of V is approximated by a scaled χ^2-distribution. The scaling factor and the degrees of freedom of the χ^2-distribution is found by matching the first two moments of V and the approximate distribution. In the following it will be utilized that $E(\chi^2_m)=m$ and Var$(\chi^2_m)=2m$ (Dobson, 2002; Johnson et al., 1994).

First, the distribution of V is approximated with a scaled χ^2-distribution of the form

$$V\dot{\sim}c\chi^2(\nu). \tag{73}$$

By equating the first two moments of V and $c\chi^2(\nu)$ we get $E(V_1+V_2)=E(c\chi^2(\nu))$, so that

$$c_1\nu_1+c_2\nu_2=c\nu. \tag{74}$$

Since V_1 and V_2 are independent (which can be shown using Fisher-Cochran's Theorem (Rao, 1973)), then $\text{Var}(V_1 + V_2) = \text{Var}(c\chi^2(\nu))$, so that

$$2c_1^2\nu_1 + 2c_2^2\nu_2 = 2c^2\nu. \tag{75}$$

The scaling factor, c, is found by inserting (74) into (75)

$$2c_1^2\nu_1 + 2c_2^2\nu_2 = 2c(c_1\nu_1 + c_2\nu_2) \tag{76}$$

so that

$$c = \frac{c_1^2\nu_1 + c_2^2\nu_2}{c_1\nu_1 + c_2\nu_2}. \tag{77}$$

The degrees of freedom, ν, is obtained by inserting (77) in (74) and rearranging, so that

$$\nu = \frac{(c_1\nu_1 + c_2\nu_2)^2}{c_1^2\nu_1 + c_2^2\nu_2}. \tag{78}$$

References

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics 25*(2), 290–302.

Bright, J., M. Aylott, S. Bate, H. Geys, P. Jarvis, J. Saul, and R. Vonk (2011). Recommendations on the statistical analysis of the comet assay. *Pharmaceutical Statistics 10*, 485–493.

Dobson, A. J. (2002). *An introduction to generalized linear models* (second ed.). Texts in Statistical Science. USA: Chapman and Hall/CRC.

Duez, P., G. Dehon, A. Kumps, and J. Dubois (2003). Statistics of the comet assay: a key to discriminate between genotoxic effects. *Mutagenesis 18*(2), 159–166.

Ejchart, A. and N. Sadlej-Sosnowska (2003). Statistical evaluation and comparison of comet assay results. *Mutation Research 534*, 85–92.

Hansen, M. K., A. K. Sharma, M. Dybdahl, J. Boberg, and M. Kulahci (2014). *In vivo* comet assay - statistical analysis and power calculations of mice testicular cells. Submitted.

Hartmann, A., E. Agurell, C. Beevers, S. brendler Schwaab, B. Burlinson, P. Clay, A. Collins, A. Smith, G. Speit, V. Thybaud, and R. R. Tice (2003). Recommendations for conducting the *in vivo* alkaline Comet assay. *Mutagenesis 18*(1), 45–51.

Hogg, R. V., J. W. McKean, and A. T. Craig (2005). *Introduction to Mathematical Statistics* (sixth ed.). USA: Prentice Hall.

Jeggo, P. A. and M. Löbrich (2007). Dna double-strand breaks: their cellular and clinical impact? *Oncogene 26*, 7717–7719.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1994). *Continuous univariate distributions* (second ed.), Volume 1 of *Wiley series in probability and mathematical statistics*. John Wiley & Sons.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1995). *Continuous univariate distributions* (second ed.), Volume 2 of *Wiley series in probability and mathematical statistics*. John Wiley & Sons.

Johnson, R., J. Freund, and I. Miller (2010). *Probability and Statistics for Engineers* (eight ed.). Prentice Hall.

Kumaravel, T. S. and A. N. Jha (2006). Reliable Comet assay measurements for detecting dna damage induced by ionising radiation and chemicals. *Mutation Research 605*, 7–16.

Lovell, D. P. and T. Omori (2008). Statistical issues in the use of the comet assay. *Mutagenesis 23*(3), 171–182.

Lovell, D. P., G. Thomas, and R. Dubow (1999). Issues related to the experimental design and subsequent statistical analysis of in vivo and in vitro comet studies. *Teratogenesis, Carcinogenesis, and Mutagenesis 19*, 109–119.

Malta, L. G., F. G. Ghiraldini, R. Reis, M. do Vale Oliveira, L. B. Silva, and G. M. Pastore (2012). In vivo analysis of antigenotoxic and antimutagenic properties of two brazilian cerrado fruits and the identification of phenolic phytochemicals. *Food Research International 49*, 604–611.

Montgomery, D. C. (2005). *Design and Analysis of Experiments* (sixth ed.). USA: John Wiley & Sons, Inc.

Olive, P. L., J. P. Banáth, and R. E. Durand (1990). Heterogeneity in radiation-induced dna damage and repair in tumor and normal cells measured using the "Comet" assay. *Radiation Research 122*(1), 86–94.

Pesarini, J. R., P. T. Zaninetti, M. O. Mauro, C. M. Carreira, J. B. Dichi, L. R. Ribeiro, M. S. Mantovani, and R. J. Oliveira (2013). Antimutagenic and anticarcinogenic effects of wheat bran *in vivo*. *Genetics and Molecular Research 12*(2), 1646–1659.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications* (second ed.). Wiley series in probability and mathematical statistics. USA: John Wiley & Sons.

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika 6*(5), 309–316.

Verde, P. E., L. A. Geracitano, L. L. Amado, C. E. Rosa, A. Bianchini, and J. M. Monserrat (2006). Application of public-domain statistical analysis software for evaluation and comparison of comet assay data. *Mutation Research 604*, 71–82.

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika 29*(3/4), 350–362.

Wiklund, S. J. and E. Agurell (2003). Aspects of design and statistical analysis in the comet assay. *Mutagenesis 18*(2), 167–175.

Zan, M. A., A. B. F. Ferraz, M. F. Richter, J. N. Picada, H. H. R. de Andrade, M. Lehmann, R. R. Dihl, E. Nunes, J. Semedo, and J. D. Silva (2013). *In vivo* genotoxicity evaluation of an artichoke (*Cynara scolymus* l.) aqueous extract. *Journal of Food Science 78*(2), T367–T371.